

Ansible Hue TDP

Currently the roles only supports the deployment of HA, SSL-enabled, Kerberos authenticated Hadoop clusters.

Prerequisites

- `java-1.8.0-openjdk` and `krb5-workstation` installed on all nodes
- Hue release (`hue_dist_file` role variable) file available in `files`
- Groups `hue_server` defined in the Ansible inventory
- Certificate files `{{ fqdn }}.key` and `{{ fqdn }}.pem` for every node available in `files`
- Certificate of the CA available as `root.pem` in `files`
- Admin access to a KDC with the `realm`, `kadmin_principal` and `kadmin_password` role vars provided
- A `krb5.conf` file with this KDC information must be available at `files/krb5.conf`
- The hue tarball must be present in the files directory. The tested version in [hue-4.10.0.tgz](#).

Notes:

- The first time you access the hue web-ui, you will have to create a hue admin user.
- Upon a successful deployment of hue, the default web-ui url is `https://<_hue_server_host>:<hue_port>`
- `yarn_rm` and `hdfs_nn` *must* be defined in the ansible hosts file
- Errors resembling `The error was: ansible.errors.AnsibleUndefinedVariable: 'dict object' has no attribute 'ats'` are often related to a missing entry in the ansible hosts file.

Example

The following hosts file and playbook are given as examples.

A ranger synched admin user `hue` with password `hue-user123` is already deployed by this role (but you can't log in as this user until your 2nd login as by default the 1st login creates a new Hue user).

Host file

```
[hue_server]
tdp-master-1
```

After deploying you will need to give the hue user in ranger access hbase tables in order to use the phoenix notebook.

Available playbooks

TODO

- Implement automatic failover (currently automates hue.ini config only on installation)
- Parameterize many of the hue.ini hardcoded values
- Create python env for running hue server with correct dependencies (especially psycog2)
- Configure hbase client
- Fix 'YARN RM returned a failed response: No connection adapters were found' to browse yarn jobs created by Hive
- Fix `Cannot access: /user/tdp_user.HTTPSConnectionPool(host='master-01.tdp', port=8020): Max retries exceeded with url: /webhdfs/v1/user/tdp_user?op=GETFILESTATUS&user.name=hue&doas=tdp_user (Caused by SSLError(SSLEOFError(8, u'EOF occurred in violation of protocol (_ssl.c:618)')))` to view hdfs using tdp_user
- Auto add hue admin user on deployment